

**UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF LOUISIANA**

**IN RE: ACTOS (PIOGLITAZONE)
PRODUCTS LIABILITY
LITIGATION**

MDL No. 6:11-md-2299

JUDGE DOHERTY

MAGISTRATE JUDGE HANNA

This Document Applies to:

ALL CASES

**PLAINTIFFS' STEERING COMMITTEE'S MOTION TO COMPEL PRODUCTION OF
DOCUMENTS AND MEMORANDUM IN SUPPORT**

The Plaintiffs' Steering Committee ("PSC") respectfully moves this Court pursuant to Fed. R. Civ. P. 37 for entry of an order compelling Defendants TAKEDA PHARMACEUTICALS U.S.A., INC. (F/K/A TAKEDA PHARMACEUTICALS NORTH AMERICA, INC.), TAKEDA PHARMACEUTICALS AMERICA, INC., TAKEDA PHARMACEUTICALS LLC, TAKEDA PHARMACEUTICALS, INTERNATIONAL INC., TAKEDA GLOBAL, RESEARCH AND DEVELOPMENT CENTER INC.), TAKEDA CALIFORNIA INC (F/K/A TAKEDA SAN DIEGO, INC., and TAKEDA PHARMACEUTICALS COMPANY LIMITED, (hereinafter referred to collectively as "Takeda") to produce further documents as set forth in this Court's Case Management Order: Protocol Relating to the Production of Electronically Stored Information, Doc. 1539, entered on July 27, 2012. In further support of this Motion, the PSC submits and incorporates the declarations of David D. Lewis, Ph.D., Exh. 1, and Douglas Forrest, Exh. 2.

BACKGROUND

On July 27, 2012 after many weeks of negotiation and court involvement, the Court entered Case Management Order Protocol Relating to the Production of Electronically Stored Information. Dkt. 1539, attached as Exhibit 3.

During the negotiations and through discussions with the Court, the Plaintiffs' Steering Committee suggested that a better mousetrap existed for finding relevant responsive documents, Predictive Coding. Predictive Coding allows for a more efficient system for finding relevant documents, and published peer-reviewed literature has shown that Predictive Coding results in lower costs and more efficient discovery processes than traditional search term methods.

More importantly, peer-reviewed literature has shown that traditional search methodology typically only resulted in the production of about 20 to 25% of relevant documents held and maintained by the defendants in their files. The same literature shows that the use of predictive coding in advanced document analytics could result in the production of over 80% of relevant documents contained within a defendant's files, and is less burdensome and more cost efficient. A more detailed description of Predictive Coding is found in the Declaration of David Lewis at paragraphs 10 to 16. Exh. 1.

Section E of the Court's Order called for the Parties to apply the use of advanced analytics as a document identification mechanism for the review and production of data and documents, as agreed by the Parties. The parties agreed to use Defendant's vendor, Epiq Systems, which used the Equivio Relevancy software for its Predictive Coding Analysis. The original order called for a proof of concept that would be applied to four key custodians identified by the parties, as well as key regulatory documents held by the company. See Order Dated July 27, 2012. Under section E.5., a document assessment phase required representatives

from the Plaintiffs Steering Committee and the Defendants' counsel to review documents to create a "control set" against which the predictive coding mechanism could be evaluated. The parties reviewed over 1,000 documents in order to arrive at a control set of approximately 385 documents. The overall richness of the control set was approximately 36% (385/1073). This meant that about 36% of the documents identified randomly by the predictive coding system in the collection of custodians were responsive and relevant to the litigation. The parties had actually agreed that certain documents, which either could not be evaluated or for which agreement was impossible, would be designated as "Skip" documents. The calculated richness of the collection through this assessment phase was approximately 40% after removing the "skip" documents.

After creation of the "control set", there was a training phase in which lawyers from the PSC and defendants participated, including Andy Birchfield, Stephanie O'Connor, and Neil Overholtz from the steering committee, and Sherry Knutson, James Mizgala, and Julie Flaming from the Defendants. This training phase was held at the offices of Nelson Mullins either in Columbia, South Carolina or Atlanta, Georgia, as had the original assessment phase. During this training phase, the attorneys reviewed "training sets" of 40 documents in which the computer attempted to learn its ability to predict whether or not a document was relevant. The parties reviewed almost 30 sets of training phase documents before the system was able to reach what is known as "stability" as verified by the Defendants' vendor.

After reaching stability, the Equivio predictive coding software assigned predictive "relevancy scores" to each document in the collection. This score is not a measure of how "relevant" a document is, or how useful a document may be, but instead represents the

software's confidence in percentage points that the document is either relevant or not. See Exh. 1, Declaration of David Lewis at para 10.

Following the training phase, the parties agreed that sampling would be done in which samples of various score points, including 20, 50, 75 and 90, would be provided to both parties by the vendor so that an analysis could be done as to the effectiveness of the predictive coding so far. After several weeks, the samples were provided and reviewed by both parties. The parties agreed, with the Special Masters approval, to extend the proof of concept beyond the original four custodians and regulatory documents to include the full set of twenty-nine (29) original custodians identified in section D of the case management order regarding the production of ESI dated July 27, 2012.

Due to the extension of the proof of concept, an additional training phase was required in order for the system to again reach "stability." This phase required review of additional "training sets" of documents by the PSC lawyers and lawyers for the Defendants until the system again reached "stability". Again the parties agreed to review samples of various score points including 20, 50, 75, and 90.

After reviewing the samples, the parties agreed to use a "seed set" of documents as contemplated by Section 2 of the court order of July 27, 2012 as a supplement to the training of the Equivio predictive coding system. The PSC provided several hundred documents broken out by key functional areas of a pharmaceutical company (marketing, regulatory, clinical, non-clinical, Pharmacovigilance), for the applicable "seed set". Counsel for defendant reviewed the documents selected by plaintiffs to be included for the seed set, and recommended several documents of their own. The parties then were able to agree on a final seeding collection. The purpose of the seed set was to provide the predictive coding system a set of documents which the

parties agreed were relevant and responsive, in an effort to improve the predictive coding system's algorithm for predicting and recognizing whether or not a document is relevant.

The Defendant's vendor's software system applied the seed set to the training process and was able to generate new predictive relevancy scores to the entire collection of documents subject to the predictive coding process, i.e., the twenty-nine (29) original custodians listed in the order along with the regulatory documents. Again, the parties agreed to review samples at 20, 50, 75, and 90 relevancy score points. The sampling allowed the parties to gauge the "precision" of the Predictive Coding software at this phase. "Precision" refers to the percentage of relevant responsive documents for any particular predictive relevance score point or range. Based on the review, the PSC was satisfied that the predictive coding process was working in a very efficient manner as their review of the sample at the 90 score point resulted in over 85% of the documents being responsive and relevant, or at least an 85% precision rate. Review of the samples at scores of 75 and 50 were also within precision estimates which were acceptable.

The Plaintiffs Steering Committee requested that predictive coding documents begin being produced immediately. At that time, and in order to avoid unnecessary and lengthy negotiations, the PSC agreed the Defendants would begin producing previously unproduced documents which carried a predictive coding score of 95 or greater.

On May 15th, the defendants produced documents for 31 of custodians from which the predictive coding software had scored the document as having a 95% confidence rating that the document contained relevant information. The PSC immediately engaged in a review of these documents. Following this review, and being pleased with the results generated through the predictive coding process of production of over 4000 documents and 16,000 pages that otherwise would not have been produced but for predictive coding, the PSC indicated to the Special

Masters their desire to request additional productions and lower score points on a rolling basis. See Exh. 1, Declaration of David Lewis at p. 4.

On May 24th the Plaintiffs' Steering Committee requested that production begin immediately at score points of 90 and above, then rolling down to 52 and above over the next several weeks. See Exh. 1, Declaration of David Lewis at p. 3. Scores of 52 had been identified by the predictive coding software as resulting in a recall rate of over 80%. Exh. 1, Declaration of David Lewis at paragraph 51.

The recall rate is a computer-generated projection of the number of relevant documents that likely will be produced by the review and production of documents above a certain score point. The predictive coding computer derives this projected percentage by examining the original 385 relevant documents identified as the "control set" and evaluating whether they would be produced at the various score points. The projected recall rate as called for by the Order Dated July 27, 2012 is a very accurate projection, as the parties agreed to a control set over five times larger than the standard control set used by Epiq and Equivio. This larger control set allows for greater statistical accuracy.

As described in the affidavit of David Lewis, plaintiffs' proposal with a recall rate of 80% is effective and provides a statistical validation of that effectiveness in the use of predictive coding software. See Ex. 1, Lewis Declaration at p.8.

On May 28th, 2013 Heyward Bonyata responded on behalf of the defendants that the defendants were opposed to production of documents below the 95 score point stating they would only be interested in further discussion of such documents if defendants could use a "search filter" on top of the documents identified by predictive coding. At a regular weekly discovery conference held with the Special Masters, the PSC and Defendants presented their

positions regarding the need to move forward with predictive coding productions, and defendants raised objection to the production of documents at scores below 95 without the addition of a filter which would use search methodology, arguing that the remaining documents not previously produced and not previously reviewed had a precision rate lower than 40%; and also arguing that reviewing such documents when precision rates were that low was overly burdensome without the addition of a search filter. While the PSC agreed to consider such request, and engaged in meet and confers to address the issue, the PSC also requested additional statistical information, so that the parties could be on equal footing with respect to the evaluation of the “precision” of the predictive coding system. The PSC believed that the Defendant’s definition of “precision” at this stage was fatally flawed, in that it failed to account for the documents that had already been reviewed and produced at those score levels; which when considered, actually demonstrated that the predictive coding software’s precision rate was very high.

The PSC requested the Special Masters require the Defendants provide this data to plaintiffs over the next several days. After receipt of the preliminary information regarding the total number of documents in the collection for each of the score points from 0 to 100 in 10% increments, as well as information regarding the number of documents already produced in those increments as well as the number of documents that had not yet been reviewed at those score points, the PSC determined through consultation with the PSC’s expert that the need for a search filter on top of the predictive coding process was unnecessary and in direct conflict with the purposes of the application of the predictive coding software as contemplated by the Order of the Court dated July 27, 2012.

Before and during the weekly conference call on Friday, June 14, 2013, the Plaintiffs’ Steering Committee indicated that they were not inclined to agree to any type of search filter to

be applied on top of the documents identified by the predictive coding software before the production. Despite many multiple meet and confers regarding this issue, and the assistance of the Special Masters, the parties have reached an impasse as to the further production of documents identified by the predictive coding software and, therefore; file this motion to compel production of predictive coding documents pursuant to the instructions of Special Masters Russo and Rodriguez.

ARGUMENT

Fed. R. Civ. P. 26(b)(1) requires a party to produce requested information that is “reasonably calculated to lead to the discovery of admissible evidence.” *Coughlin v. Lee*, 946 F.2d 1152, 1159 (5th Cir. 1991) (quoting Fed. R. Civ. P. 26(b)(1)); *Amos v. Jackson*, 2006 WL 3087244 at *3 (W.D.La. Oct. 25, 2006) (not reported) (Doherty, J.). The same rule also “provides that a party must provide information and documents it possesses, regardless of who else possesses that information.” *U.S. ex rel. Mallavarapu v. Acadiana Cardiology, LLC*, 2012 WL 369896 at *5 (W.D.La. Feb. 3, 2012) (Hanna, M.J.).

Relevancy is broadly construed as “courts have held “information is relevant if it encompass[es] any other matter that bears on, or that reasonably could lead to other matter that could bear on, any issue that is or may be in the case.” *Coughlin*, 946 F.2d at 1159 (internal quotations omitted); *see also Merrill v. Waffle House, Inc.*, 227 F.R.D. 467, 470 (N.D.Tex. 2005) (holding information “should be considered relevant if there is any possibility that the information sought may be relevant to the claim or defense of any party.”) (internal quotations omitted). Indeed, the standard is not that relevant documents be produced; rather documents that are not irrelevant are to be produced. *Alonso v. Agrigenetics, Inc.*, 2004 WL 2668801, at *3

(S.D.Tex. 2004). Finally, the burden is on the party resisting production “to establish the lack of relevancy....” *Merill*, 227 F.R.D. at 470-71.

Here, the parties have jointly trained the predictive coding software as part of the collaborative process set forth in this Court’s Case Management Order on Electronically Stored Information. The predictive coding strategy proposed by the PSC makes effective use of predictive coding, while also making use of the review work already performed by Takeda. The PSC has suggested a goal of obtaining 80% (not 100%) of relevant documents. In order to achieve a recall rate of only 80% of relevant documents, production of documents with a predictive coding relevancy score of approximately 52 or greater would be required. To that end, the PSC has offered starting at the 90% predictive coding relevancy score, with the parties working towards an approximately 50% relevancy score on a rolling basis over several weeks. It should be noted, that Takeda will not have to review or produce documents previously reviewed and produced from these custodians' files.

Yet, Takeda has resisted the PSC’s proposal that is reasonably geared to obtain 80% of the production. The PSC is not even attempting to obtain 100% of relevant production. Takeda has further attempted to confound the purpose of predictive coding by adding search filters. Predictive coding does not require these search term filters. Takeda’s efforts attempt to limit the production of relevant documents despite express Fifth Circuit language that relevancy is broadly construed and any relevant documents must be produced. Takeda’s position that predictive coding has been ineffective is not supported by their data. Exh. 1, Lewis Declaration, paragraphs 29 through 41. Moreover, Takeda’s proposal for selecting documents for review cannot be relied upon, as the proposal would provide few additional responsive documents beyond those already. Exh. 1, Lewis Declaration, paragraphs 42 through 50. Plaintiffs have instead proposed an

effective use of predictive coding that would provide a statistical validation of its effectiveness. Exh. 1, Lewis Declaration, paragraphs 51 through 53.

Essentially, to analogize, the parties , along with, the input of the Court and the Court appointed Special Masters, worked diligently to create a very sophisticated tomato picker that would go through all of the tomato plants, and pick all the tomatoes, the “HAL” Tomato Picker. HAL would then sort them, and even would deliver the tomatoes, with the freshest, ripest tomatoes coming off first. The time and expense of building this “HAL” tomato picker has been well documented and described. The parties put in all of the necessary work to fine tune the system. But in the meantime, the Defendants, because they had a delivery expected in Illinois, used manual tomato pickers, to “search” through and find fresh ripe tomatoes to deliver. The manual pickers left a lot of good ripe tomatoes still on the vine though.

As it turns out, the new “HAL” picker would have picked most of the ones the manual pickers picked, and would have also found the ones hiding under leaves. Now that what’s left on the vines are either green tomatoes, or rotten tomatoes, and maybe fewer super fresh red ripe tomatoes than before the manual pickers took them all, the Defendants want to say it’s useless to use the fancy HAL picker we built, arguing that because they picked most of the good stuff manually, the cost of using the machine to go through what’s left isn’t worth it, as too many rotten tomatoes will get through, and will have to be culled out. The Defendant’s argument misses the point. We know there are still tons of good tomatoes, many of which would never be found by a regular search, or even a new search, and the only way to get them is to use our HAL picker as it was designed.

Pursuant to this Court’s Case Management Order regarding Electronically Stored Information, Plaintiffs are entitled to the requested discovery.

CONCLUSION

For the foregoing reasons, this Court should grant the PSC's Motion to Compel in its entirety and order Takeda to produce non-privileged documents pursuant to this Court's July 27, 2012 Case Management Order.

LR 37.1 Statement

As required by this Court's Local Rule 37.1, counsel for the Plaintiffs' Steering Committee hereby certifies that said counsel has conferred with counsel for Defendants for the purpose of amicably resolving the issues. Despite many multiple meet and confers regarding this issue, and the assistance of the Special Masters, the parties have reached an impasse as to the further production of documents identified by the predictive coding software and, therefore; file this motion to compel production of predictive coding documents pursuant to the instructions of Special Masters Russo and Rodriguez.

Dated: June 17, 2013

By: /s/ Neil D. Overholtz
Neil D. Overholtz, Esq.
On behalf of the Plaintiffs Steering Committee
Aylstock, Witkin, Kreis, & Overholtz, PLLC
17 East Main Street
Suite 200
Pensacola, Florida 32502
(850) 202-1010
(850) 916-7449
noverholtz@awkolaw.com

CERTIFICATE OF SERVICE

I certify that on 17th day of June, 2013, I electronically filed the foregoing with the Clerk of the Court by using the CM/ECF system, which will send a notice of electronically filing to all parties of record by operation of the Court's electronic filing system.

By: /s/ Neil D. Overholtz